Identification of developmental trajectory classes

Sijbrandij, Jitske J.; Hoekstra, Tialda; Almansa, Josué; Reijneveld, Sijmen A.; Bültmann, Ute

Published in:
Advances in Life Course Research

DOI:
10.1016/j.alcr.2019.04.018

Identification of developmental trajectory classes: Comparing three latent class methods using simulated and real data

Authors

Jitske J. Sijbrandij[a] (corresponding author) (j.j.sijbrandij@umcg.nl)

Tialda Hoekstra[a] (t.hoekstra@umcg.nl)

Josué Almansa[a] (j.almansa.ortiz@umcg.nl)

Sijmen A. Reijneveld[a] (s.a.reijneveld@umcg.nl)

Ute Bültmann[a] (u.bultmann@umcg.nl)

Affiliations and addresses

[a] Department of Health Sciences, Community and Occupational Medicine

University of Groningen, University Medical Center Groningen, Groningen, the Netherlands

Antonius Deusinglaan 1, Building 3217, 9713 AV Groningen

Abstract

Introduction Several statistical methods are available to identify developmental trajectory classes, but it is unclear which method is most suitable. The aim of this study was to determine whether latent class analysis, latent class growth analysis or growth mixture modeling was most appropriate for identifying developmental trajectory classes.

Methods We compared the three methods in a simulation study in several scenarios, which varied regarding e.g. sample size and degree of separation between classes. The simulation study was replicated with a real data example concerning anxiety/depression symptoms measured over 6 time points in the Tracking Adolescent Individuals' Lives Survey (TRAILS, N= 2,227)

Results Growth mixture modeling was least biased or equally biased compared to latent class analysis and latent class growth analysis in all scenarios. In TRAILS, the shapes of the trajectories were rather similar over the three methods, but class sizes differed slightly. A 4-class growth mixture model performed best, based on several fit indices, interpretability and clinical relevance.

Conclusions Growth mixture modeling seems most suitable to identify developmental trajectory classes.

Keywords

Developmental trajectory; Growth Mixture Modeling; Latent class analysis; Latent class growth analysis; Longitudinal data analysis; Monte Carlo simulation study

Manuscript

1. Introduction

Longitudinal data allows researchers to assess the development of individuals over time (Twisk 2013) and to answer questions such as "How do mental health problems (Wickrama et al. 2008; Veldman et al. 2017), drinking behavior (Virtanen et al. 2015) or smoking behavior (Daw et al. 2017; Patel et al. 2017) of adolescents develop into adulthood?". However, the analysis of longitudinal data is often complicated by a low prevalence of disease or risk behaviors, skewed distributions and categorical data (Feldman et al. 2009).

Classifying individuals based on their development over time can be an intuitive tool to illustrate heterogeneity within a population and can offer a solution to the statistical issues mentioned above. Moreover, questions such as "What different classes of development are identified and how do these differ from one another?'' can be answered.

Developmental trajectory classes can be estimated for statistical or theoretical reasons, which leads to different interpretations of the developmental trajectory classes. When estimating classes for statistical, i.e. practical, reasons, as mentioned above , making groups is a way to summarize the data. Both in the statistical and theoretical interpretation, LCA, LCGA and GMM estimate an individual's probability to belong to a certain class, they do not simply assign individuals to classes. The classes are not considered to be actual observed groups, but estimated latent groups (Nagin and Tremblay 2005; Nagin and Odgers 2010). Several authors have therefore warned to not reify the classes (Nagin and Tremblay 2005; Nagin and Odgers 2010; Sher et al. 2012).

Gangestad and Snyder (1987) have a different philosophy and consider the classes to be more theoretical concepts. They consider classes which aim to summarize the data, to be oversimplifications (Gangestad and Snyder 1985). The authors have two requirements for classes. First, classes (not necessarily being developmental trajectory classes) are not merely summaries of the data, but should have a causal origin. Second, "class variables explicated as latent constructs either really exist or they do not, and class variables explicated in these terms either really exercise influence upon phenotypic characteristics or they do not.

To date, it remains unknown which statistical method is most suitable for classifying individuals based on their development over time (also referred to as developmental trajectory classes). Frequently applied methods are latent class analysis (Lanza and Collins 2006) (LCA, also referred to as latent profile analysis), latent class growth analysis (Nagin 1999) (LCGA, also known as group-based trajectory modeling (Nagin and Odgers 2010)) and growth mixture modeling (Muthen and Shedden 1999) (GMM, also known as latent class mixture modeling (Hoekstra 2013)) (Figure 1). Latent class analysis estimates means and variances for each time point per class and does not take the longitudinal nature of the data into account. Latent class growth analysis estimates a mean intercept and slope per class. Growth mixture modeling is an extension of LCGA, which allows individuals to vary around the class' intercept and slope (similar to random effect in multilevel analysis). Berlin et al. wrote an easy to comprehend, non-technical introduction to LCA (Berlin et al. 2014b), LCGA and GMM (Berlin et al. 2014a).

Although the use of LCA, LCGA or GMM can result in different outcomes (Morin et al. 2011; Kooken et al. 2018), it is unknown which of these three methods performs best. The relative performances of LCA, LCGA and GMM has previously been assessed (Twisk and Hoekstra 2012; Martin and von Oertzen 2015; Diallo et al. 2016; Davies et al. 2017), but the three methods have not yet been compared together in one simulation study, nor have they been compared in terms of bias. Moreover, previous studies did not include a real data example.

1A) Latent Class Analysis 1B) Latent Class Growth Analysis 1C) Growth Mixture Modeling

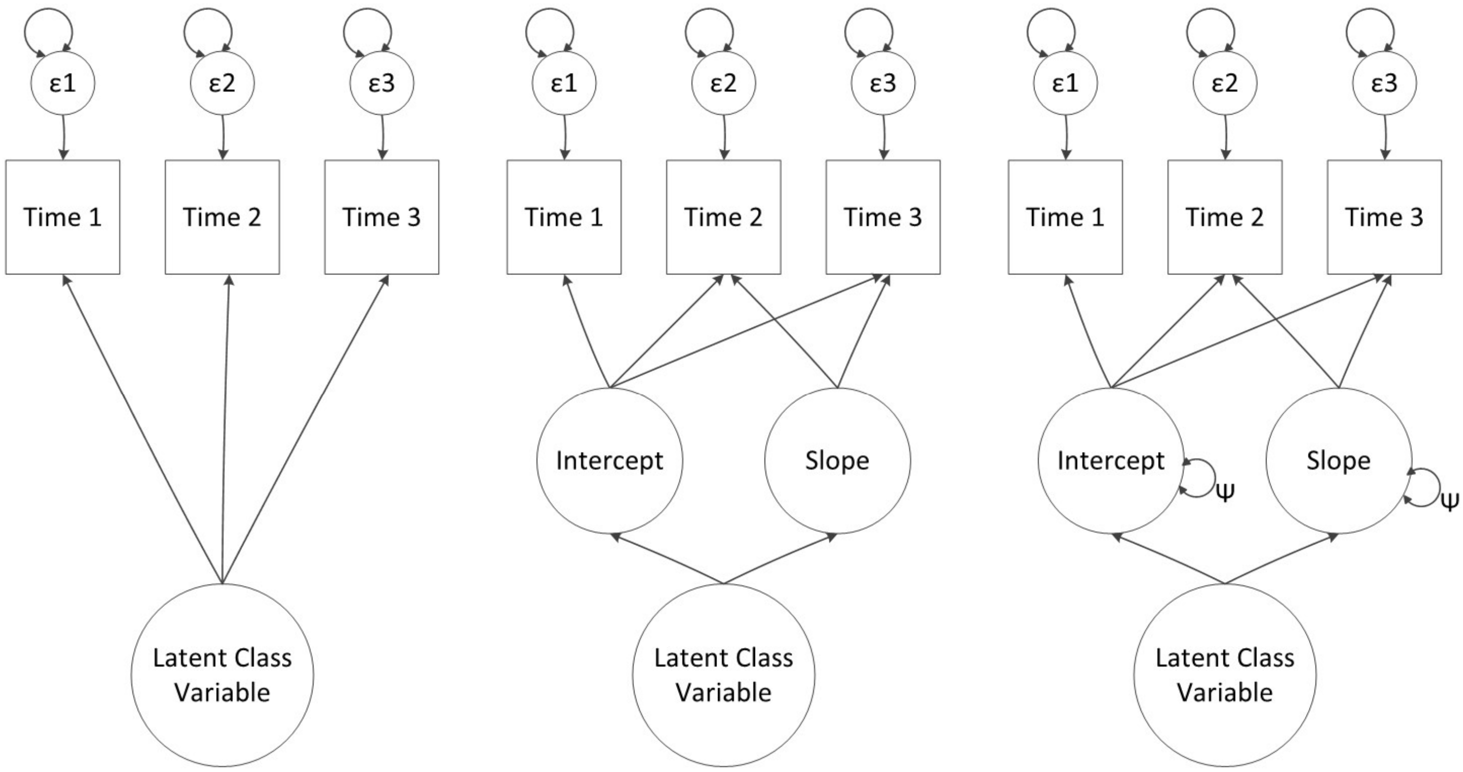

Figure 1 Diagram of Latent Class Analysis, Latent Class Growth Analysis and Growth Mixture Modeling

Figure note: The circular arrows above the epsilons represent the residual variances. The circular arrows accompanied with the letter psi represent the variances accounted for in the model.

A real data example can help translating the implications of the simulation study and can hereby improve understandability. The few available comparison studies are limited in that they mainly focused on identifying the simulated number of classes (Diallo et al. 2016). A model that does not extract the simulated number of classes may be a close representation of reality (e.g. when similar classes are merged), while a model that extracts the simulated number of classes is not (e.g. when almost empty classes are estimated) (Martin and von Oertzen 2015). Earlier comparisons did not employ the degree of bias, i.e. the difference between the true and estimated parameter values of the methods, as outcome (Twisk and Hoekstra 2012; Diallo et al. 2016). Bias is directly related to the accuracy of methods (Burton et al. 2006). Moreover, relevant scenarios, e.g. a quadratic growth trend, were not covered in previous research (Martin and von Oertzen 2015; Diallo et al. 2016; Davies et al. 2017). The lack of a comprehensive simulation study hinders researchers in making informed decisions about which method to apply, which may result in biased conclusions.

The aim of this study was to assess which method, LCA, LCGA or GMM, is most appropriate to identify developmental trajectory classes and under which circumstances. We addressed this aim in a simulation study and in a real data example.

2. Simulation study

The simulation study consisted of four steps: 1) a systematic literature search to provide realistic input values; 2) data generation; 3) analysis of generated data; and 4) assessment of the outcomes, i.e. class recovery, bias and model selection.

2.1 Methods simulation study

2.1.1 Systematic literature search

As suggested by Paxton et al. (Paxton et al. 2001), we performed a systematic literature search to provide realistic input values for the simulation study on sample size, number of time points, number of underlying classes, relative class sizes and separation between classes (Supplementary Material 1). We searched the top 25% journals of the Web of Science category *Public, Environmental & Occupation Health*, based on the Social Science Citation Index and Science Citation Index of the last three years (2014-2016). Studies published in 2016-2017 that applied LCA, LCGA or GMM to longitudinal data were selected.

2.1.2 Data generation

First, a basic scenario was generated, corresponding to the median values of the literature search. This resulted in a scenario with 2,000 participants, five time points and a trajectory variable of four classes with linear growth over time. The trajectory variable could entail a continuous variable such as anxiety/depression symptoms (as in the real data example) or low back pain (as taken from the literature search). A high score would then entail high symptoms and a low score low symptoms. The four classes were defined as follows: the first class contained 60% of the sample (labeled ''low stable"), the second class contained 20% ("low increasing") and the third and fourth class both contained 10% of the sample ("high decreasing" and "high stable", respectively). The separation between the classes was set at 1 standard deviation (SD), i.e. the mean value at the first time point differed 1 SD for adjacent classes.

Second, we defined 13 different scenarios to assess their effects on the relative performance of LCA, LCGA and GMM (Supplementary Material 3, Table 6). The scenarios were created by taking the corresponding 25^{th} and 75^{th} percentile of the results of the systematic literature search. The scenarios differed regarding sample size (1,000 or 8,000), number of time points (four or seven), number of underlying classes (three or five), relative class size (largest class containing 40% or 75% of the sample) and separation between classes (0.5, 2 or 3 SDs). Additionally, a sample size of 200 was included, because we were particularly interested in the performance of the methods with smaller sample sizes. Lastly, we generated a scenario with quadratic growth.

We generated the data according to a GMM and an LCGA model. We present the results for data generated according to GMM; results for LCGA generated data are presented when these differed from the GMM results. The variances and residual variances were constrained to be equal across classes and over time, and the intercept and slope parameters were uncorrelated. In sensitivity analyses, we generated the standard scenario, but with a correlation of -0.2 between the random intercept and random slope. For each scenario, we created 1,000 data sets in R 3.4.1 (R Core Team 2017). The R code for the simulation study is available on the Open Science Framework through the following link: https://osf.io/cqfa3/?view_only=96d06e016e904b70bd915fc0e6994920

2.1.3 Outcomes

The primary outcome was bias of the models with the same number of classes as simulated. We considered a standardized bias (bias/standard error (SE)) of more than 40% as problematic (Burton et al. 2006). The coverage indicated how often the 95% confidence interval contained the true parameter value.

In some replications, the simulated and estimated classes could not be linked to each other, because the classes were severely distorted and therefore (especially for the third class) the trajectories could not be properly distinguished. If this was the case, the replications were not included in the bias and coverage calculations. The secondary outcome was model selection, i.e. the frequency of selecting the model with the simulated number of classes over a model with one class less or more. Four fit indices were used to select the best fitting model: the BIC (Bayesian Information Criterion), AIC (Akaike Information Criterion), aBIC (sample size adjusted BIC), and LMR-LRT (Lo-Mendell-Rubin Likelihood Ratio Test). The p-value of the LMR-LRT determines whether the model significantly improves compared to a model with one class less (Lo et al. 2001). Moreover, the entropy was calculated. The entropy indicates how well classes are separated and how well individuals fit in their classes, ranging from 0 (no separation between classes) to 1 (perfect separation). Entropy is not a fit index, but is often used as such.

2.1.4 Analytical procedure

Each data set was analyzed with LCA, LCGA and GMM in Mplus version 8 (Muthén and Muthén). For LCA the means per time point were automatically calculated. To make the results of LCA comparable to LCGA and GMM, the means were converted to an intercept and slope parameter by regressing the means on time. The SE was calculated as an indicator of the variability of the estimates between replications (Burton et al. 2006). The R package *MplusAutomation* (Hallquist and Wiley 2018) was used for the communication between R and Mplus.

The number of random starting values was 100, of which 10 were used for optimization in the final stage. Because the quadratic models were more complex, 500 starting values were specified, of which 100 were used in the final stage.

2.2 Results simulation study

2.2.1 Primary outcome: Class recovery and bias

For all scenarios the class recovery was higher for GMM and the biases were consistently smaller for GMM than for LCA and LCGA (Table 1). The coverage was closest to 95% for GMM, and the SEs were usually similar over the three methods. The standardized bias was larger for LCA and LCGA than for GMM (up to 233%, 366% and 28%, respectively). In other words, for GMM, the standardized bias never exceeded the 40% cut-off.

The separation between classes influenced relative model performance the most. For small separation under LCA and LCGA, the simulated and estimated classes could not be linked, i.e. class recovery was close to 0%. For small separation under GMM, **class recovery was 60.7%** of the replications. For separation of 2 or 3 SDs, class recovery was close to 100% for all three methods and the biases were more comparable across the methods.

The sample size, number of time points, number of classes, and inclusion of a quadratic slope all affected the relative model performance in terms of class recovery, but not in terms of bias. For a sample size of 200, class recovery was 49.7% (LCA), 48.2% (LCGA) and 77.5% (GMM) of the replications. For GMM the class recovery was close to 100% for all other sample sizes, but this was not the case for LCA and LCGA. For a sample size of 1000, LCA and LCGA remained around 80% and for a sample size of 2000 they recovered the classes in 87.2% and 88.4% of the replications, respectively. For a sample size of 8000, the class recovery was comparable over methods: 99.3% (LCA), 99.7% (LCGA) and 100% (GMM) of the replications. For 7 time points LCA and LCGA the class recovery was close to 0%, while being 99.5% for GMM. When three latent classes were present, all methods attained high class recovery (ranging from 98.8-100%). This was not the case for five class solution, for which class recovery was 59.0% (LCA), 60.2% (LCGA) and 99.7% (GMM). For a quadratic slope, class recovery occurred in 9.9% (LCA), 9.8% (LCGA) and 93.6% (GMM) of the replications. When the theoretical classes were not discovered, Class 3 could not be properly distinguished from the other classes.

When the largest class contained 40% of the individuals (i.e. when relative class sizes were more equal), the biases of LCA and LCGA were smaller than when the largest class contained 60% (under the standard scenario). Moreover, when the largest class was 40% the class recovery was close to 100% for all methods. When the largest class contained 75% of the individuals, class recovery was possible in 2.1% (LCA), 2.3% (LCGA) and 99.7% (GMM) of the replications. In all other replications, the classes were so distorted that the simulated classes could not be recovered and the bias could not be calculated.

Even though the bias was smallest for GMM, warnings were also most common for GMM. The most frequent warning was due to negative variance, also known as a Heywood case (Chen et al. 2001) (results not shown).

When the data were generated according to LCGA, the bias and coverage remained small for all methods and almost all scenarios: the standardized bias remained under the 40% cut-off (up to 31%) (Supplementary Material 3, Table 7). The exception is the scenario with quadratic growth: the 40% cutoff was equally exceeded to a large extent for all three methods. For instance in class 3 the standardized bias of the intercept was over 200% for LCA, LCGA and GMM. The class recovery ranged from 89.5.1-100% (LCA), 92.0-100% (LCGA) and 82.6-100% (GMM).

Sensitivity analyses on GMM data with a correlation of -0.2 between the random intercept and random slope yielded similar findings (Supplementary Material 3). Only the numbers sometimes slightly differed.

2.2.2 Secondary outcome: model selection.

The BIC and aBIC usually selected the model with the simulated number of classes for GMM, but extracted too many classes for LCA and LCGA in most scenarios (Table 2). For all methods, the AIC often extracted too many classes. The performance of the LMR-LRT differed greatly over the scenarios, selecting the simulated number of classes up to 83.6% for GMM and never reaching 30% for LCA and LGCA. The entropy extracted too few classes for all methods and scenarios.

With the LCGA generated data, the performance of the BIC and aBIC was mostly similar for the three methods (Table 2). Only for a sample size of 200 or a class separation of 0.5 SD, LCGA selected the model with the simulated number of classes more often than LCA and GMM.

Table 1 Parameter Biases. Coverage and Standard Errors for the Data as Simulated According to the Growth Mixture

			Class 1 “Low Stable’’						Class 3 “Decreasing”						Proportion
		Class	Intercept			Slope			Intercept			Slope			correct
Model	Method	Recovery[a]	Bias	SE	Cov.	Bias	SE	Cov.	Bias	SE	Cov.	Bias	SE	Cov.	Classified[b]
Basic model	LCA	872	-0.04	0.04	0.85	-0.02	0.04	0.93	-0.25	0.20	0.56	0.12	0.12	0.86	0.87
	LCGA	884	-0.05	0.04	0.80	-0.01	0.01	0.81	-0.26	0.18	0.89	0.14	0.07	0.70	0.87
	GMM	1000	0.00	0.04	0.95	0.00	0.01	0.95	0.00	0.12	0.95	0.00	0.05	0.96	0.89
Sample size = 200	LCA	497	-0.04	0.11	0.92	-0.01	0.11	0.94	0.02	0.41	0.90	0.02	0.35	0.94	0.86
	LCGA	482	-0.04	0.10	0.94	-0.01	0.04	0.95	-0.02	0.36	0.94	0.05	0.16	0.90	0.86
	GMM	775	-0.01	0.11	0.91	0.00	0.04	0.93	0.12	0.43	0.89	-0.04	0.18	0.88	0.87
Sample size = 1.000	LCA	796	-0.04	0.05	0.90	-0.02	0.05	0.95	-0.20	0.24	0.71	0.10	0.17	0.88	0.87
	LCGA	811	-0.05	0.05	0.89	-0.01	0.02	0.91	-0.22	0.22	0.96	0.13	0.09	0.86	0.87
	GMM	999	0.00	0.05	0.94	0.00	0.02	0.94	0.00	0.17	0.93	0.00	0.08	0.94	0.88
Sample size = 8.000	LCA	993	-0.04	0.02	0.65	-0.02	0.02	0.87	-0.28	0.12	0.24	0.14	0.07	0.67	0.87
	LCGA	997	-0.05	0.02	0.33	-0.01	0.01	0.34	-0.28	0.11	0.25	0.15	0.04	0.02	0.87
	GMM	1000	0.00	0.02	0.95	0.00	0.01	0.94	0.00	0.06	0.96	0.00	0.02	0.94	0.89
4 time points	LCA	949	-0.05	0.04	0.90	-0.02	0.05	0.95	-0.21	0.22	0.82	0.08	0.19	0.94	0.85
	LCGA	965	-0.05	0.04	0.81	-0.01	0.02	0.93	-0.23	0.21	0.86	0.11	0.10	0.83	0.85
	GMM	1000	0.00	0.04	0.96	0.00	0.02	0.96	-0.01	0.16	0.95	0.00	0.08	0.95	0.86
7 time points	LCA	2	-0.04	0.08	0.50	-0.02	0.00	1.00	-0.14	0.41	0.00	0.10	0.18	0.50	0.88
	LCGA	1	-0.11		0.00	-0.01		1.00	-0.45		1.00	0.24		1.00	0.88
	GMM	995	0.00	0.03	0.96	0.00	0.01	0.94	0.00	0.08	0.95	0.00	0.03	0.94	0.92
3 classes	LCA	988	-0.04	0.04	0.84	-0.03	0.04	0.91	-0.26	0.17	0.51	0.12	0.11	0.82	0.88
	LCGA	989	-0.04	0.04	0.78	-0.02	0.01	0.81	-0.26	0.15	0.68	0.13	0.07	0.51	0.88
	GMM	1000	0.00	0.04	0.93	0.00	0.01	0.95	0.00	0.08	0.95	0.00	0.03	0.95	0.90
5 classes	LCA	590	-0.05	0.04	0.83	-0.02	0.04	0.93	-0.26	0.21	0.55	0.12	0.14	0.84	0.86
	LCGA	602	-0.05	0.04	0.80	-0.01	0.01	0.86	-0.28	0.20	0.92	0.15	0.08	0.80	0.87
	GMM	997	0.00	0.03	0.96	0.00	0.01	0.96	-0.01	0.14	0.94	0.00	0.05	0.94	0.88
.5 SD separation	LCA	0													
	LCGA	0													
	GMM	607	0.01	0.06	0.91	-0.01	0.03	0.91	0.12	0.28	0.95	-0.07	0.15	0.93	0.77
2 SD separation	LCA	1000	0.00	0.03	0.95	0.00	0.03	0.96	-0.01	0.09	0.92	-0.05	0.09	0.92	0.97
	LCGA	1000	0.00	0.02	0.94	0.00	0.01	0.95	-0.03	0.09	0.94	0.00	0.04	0.94	0.97
	GMM	1000	0.00	0.02	0.94	0.00	0.01	0.95	0.00	0.08	0.95	0.00	0.03	0.95	0.97
3 SD separation	LCA	1000	0.00	0.03	0.95	0.00	0.03	0.96	0.00	0.08	0.92	0.00	0.08	0.95	0.99
	LCGA	1000	0.00	0.02	0.95	0.00	0.01	0.95	0.00	0.07	0.94	0.01	0.03	0.93	0.99
	GMM	1000	0.00	0.02	0.95	0.00	0.01	0.95	0.00	0.07	0.94	0.00	0.03	0.94	0.99
Largest class = 40%	LCA	1000	-0.01	0.04	0.94	-0.03	0.02	0.61	-0.08	0.09	0.88	0.06	0.04	0.64	0.84
	LCGA	1000	0.00	0.05	0.91	-0.04	0.04	0.90	-0.07	0.09	0.82	0.05	0.07	0.93	0.84
	GMM	993	0.00	0.05	0.95	0.00	0.02	0.95	0.00	0.09	0.95	0.00	0.06	0.94	0.85
Largest class = 75%	LCA	21	-0.05	0.03	0.71	-0.01	0.01	0.95	-0.14	0.23	1.00	0.12	0.09	0.95	0.90
	LCGA	23	-0.05	0.03	0.74	-0.01	0.03	0.91	-0.12	0.24	0.83	0.07	0.13	0.96	0.90
	GMM	997	0.00	0.03	0.93	0.00	0.01	0.94	-0.01	0.21	0.92	0.00	0.08	0.93	0.92
Quadratic	LCA	99	-0.03	0.03	0.93	-0.03	0.04	0.96	0.23	0.17	0.70	-0.23	0.12	0.79	0.82
	LCGA	98	-0.03	0.03	0.89	-0.03	0.03	0.86	0.23	0.17	0.64	-0.23	0.11	0.49	0.82
	GMM	936	0.00	0.04	0.90	0.01	0.05	0.80	-0.02	0.24	0.92	0.02	0.28	0.95	0.85

Cov. coverage of the 95% confidence interval, *GMM* Growth Mixture Modeling, *LCA* Latent Class Analysis, *LCGA* Latent Class Growth Analysis, *SE* Standard Error

a) Only replications in which the simulated and estimated classes matched, where the simulated classes were recovered, were included in the calculation of bias, SE and coverage. In other words, if one of the estimated classes was a mixture of the estimated classes, the replication was not included. Replications were also not included if the model did not converge.

b) The proportion of individuals correctly classified was only calculated for replications where the classes were recovered , as for the other variables in this table.

Table 2 Proportion of Models With Simulated Number of Classes Correctly Identified for the GMM and LCGA Data Generations

		GMM					LCGA					Estimated parameters
	Analysis method	BIC	aBIC	AIC	LMR-LRT	Entropy	BIC	aBIC	AIC	LMR-LRT	Entropy	
Basic model	LCA	0.00	0.00	0.00	0.04	0.44	1.00	1.00	0.44	0.84	0.00	28
	LCGA	0.00	0.00	0.00	0.03	0.42	1.00	1.00	0.96	0.64	0.00	16
	GMM	1.00	0.99	0.69	0.83	0.26	1.00	0.99	0.67	0.79	0.17	19
N = 200	LCA	0.37	0.12	0.09	0.23	0.34	0.58	0.56	0.45	0.49	0.10	28
	LCGA	0.57	0.14	0.12	0.33	0.33	0.89	0.86	0.85	0.66	0.08	16
	GMM	0.23	0.54	0.52	0.26	0.27	0.57	0.67	0.63	0.47	0.25	19
N = 1,000	LCA	0.09	0.00	0.00	0.22	0.43	1.00	0.97	0.47	0.85	0.01	28
	LCGA	0.00	0.00	0.00	0.14	0.42	1.00	0.99	0.93	0.71	0.01	16
	GMM	0.99	0.96	0.68	0.85	0.24	1.00	0.96	0.66	0.83	0.17	19
N = 8,000	LCA	0.00	0.00	0.00	0.00	0.45	1.00	1.00	0.39	0.85	0.00	28
	LCGA	0.00	0.00	0.00	0.00	0.44	1.00	1.00	0.99	0.72	0.00	16
	GMM	1.00	1.00	0.71	0.82	0.25	1.00	1.00	0.68	0.80	0.09	19
4 time points	LCA	0.35	0.03	0.00	0.24	0.03	1.00	1.00	0.51	0.81	0.01	23
	LCGA	0.06	0.00	0.00	0.16	0.03	1.00	0.99	0.77	0.79	0.01	15
	GMM	1.00	0.99	0.66	0.84	0.24	1.00	0.99	0.67	0.82	0.11	18
7 time points	LCA	0.00	0.00	0.00	0.08	0.14	1.00	1.00	0.44	0.92	0.00	38
	LCGA	0.00	0.00	0.00	0.09	0.15	1.00	1.00	0.89	0.89	0.00	18
	GMM	1.00	0.99	0.70	0.83	0.29	1.00	0.98	0.72	0.80	0.20	21
3 classes	LCA	0.00	0.00	0.00	0.10	0.00	1.00	1.00	0.45	0.83	0.00	22
	LCGA	0.00	0.00	0.00	0.10	0.00	1.00	0.99	0.81	0.87	0.00	13
	GMM	1.00	0.99	0.71	0.84	0.00	1.00	0.99	0.70	0.83	0.00	16
5 classes	LCA	0.00	0.00	0.00	0.05	0.75	1.00	1.00	0.44	0.88	0.00	34
	LCGA	0.16	0.16	0.15	0.13	0.67	0.99	0.99	0.95	0.78	0.01	19
	GMM	1.00	0.98	0.66	0.77	0.07	0.99	0.98	0.67	0.71	0.02	22
.5 class separation	LCA	0.57	0.11	0.00	0.32	0.35	0.53	0.96	0.46	0.77	0.01	28
	LCGA	0.16	0.04	0.02	0.30	0.33	0.95	0.96	0.87	0.81	0.04	16
	GMM	0.03	0.29	0.55	0.29	0.07	0.20	0.60	0.66	0.46	0.06	19
2 class separation	LCA	0.00	0.00	0.00	0.01	0.00	1.00	1.00	0.50	0.82	0.00	28
	LCGA	0.00	0.00	0.00	0.01	0.00	1.00	0.99	0.78	0.81	0.00	16
	GMM	1.00	0.98	0.65	0.79	0.00	1.00	0.99	0.67	0.78	0.00	19
3 class separation	LCA	0.00	0.00	0.00	0.00	0.00	1.00	1.00	0.49	0.83	0.00	28
	LCGA	0.00	0.00	0.00	0.00	0.00	1.00	0.99	0.68	0.71	0.00	16
	GMM	1.00	0.98	0.67	0.76	0.00	1.00	0.98	0.68	0.76	0.00	19
Largest class= 0.4	LCA	0.02	0.00	0.00	0.21	0.88	1.00	1.00	0.46	0.82	0.44	28
	LCGA	0.00	0.00	0.00	0.19	0.87	1.00	0.99	0.82	0.85	0.52	16
	GMM	0.99	0.98	0.71	0.75	0.50	0.97	0.96	0.71	0.62	0.56	19
Largest class=0.75	LCA	0.00	0.00	0.00	0.11	0.23	1.00	1.00	0.47	0.88	0.00	28
	LCGA	0.00	0.00	0.00	0.07	0.22	0.98	0.98	0.90	0.81	0.01	16
	GMM	1.00	0.98	0.69	0.82	0.05	1.00	0.99	0.71	0.81	0.00	19
Quadratic	LCA	0.01	0.00	0.00	0.10	0.00	1.00	1.00	0.40	0.82	0.46	16
	LCGA	0.00	0.00	0.00	0.07	0.00	1.00	1.00	0.71	0.81	0.59	20
	GMM	0.39	0.83	0.77	0.65	0.17	0.99	0.99	0.76	0.73	0.46	23

AIC Akaike Information Criterion, *aBIC* sample size adjusted BIC, *BIC* Bayesian Information Criterion, *GMM* Growth Mixture Modeling, *LCA* Latent Class Analysis, *LCGA* Latent Class Growth Analysis, *LMR-LRT* Lo-Mendell-Rubin Likelihood Ratio Test

3 Real data example: The TRAILS Cohort

The application to the real data consisted of three steps: 1) analysis of the data; 2) selection of the final models and 3) comparison and interpretation of final models.

3.1 Methods

The use of LCA, LCGA and GMM to identify developmental trajectory classes was illustrated with the TRacking Adolescents' Individual Lives Survey (TRAILS) cohort (Huisman et al. 2008; Oldehinkel et al. 2015). The TRAILS study assesses mental, physical and social development of adolescents in the Netherlands. The TRAILS cohort consists of children born in five municipalities in the north of the Netherlands, between October 1st 1989 and September 30th 1991. In total, 2,230 children, with a mean age of 11.1 years (SD = 0.56), participated in the study at baseline in 2000/2001, 51% of which were girls. Data were collected every two to three years, at in total six time points, with a mean follow-up of 14.5 years (SD = 0.48). The mean age at the second measurement was 13.6 (SD = 0.53), 16.3 at the third measurement (SD = 0.71), 19.1 at the fourth measurement and 22.3 at the fifth measurement wave (SD = 0.60). At the sixth measurement wave, 1,617 individuals participated (72.7% of the original sample), with a mean age of 25.7 years (SD = 0.60), 53% of which were girls. Further details on the TRAILS cohort have been reported elsewhere (Huisman et al. 2008; Oldehinkel et al. 2015). The TRAILS data are available for public use and can be requested at www.trails.nl.

3.1.1 Measures

Anxiety/depression symptoms were assessed with the Youth Self-Report (Achenback and Rescorla 2001) for time points 1-3 (consisting of 13 items) and the Adult Self-Report (Achenbach and Rescorla 2003) for time points 4-6 (consisting of 18 items). The item scores ranged from 0 ("Not true") to 2

(“Very true or often true“).To make scores on the two scales comparable, we transformed the scales such that they both ranged from 0 to 20. To do so, we first calculated the mean score both scales (mean score ranging from 0 to 2). Then, we multiplied this score by ten, resulting in a scale from 0 (no symptoms) to 20 (maximum symptoms). Individuals’ measurements on time points were included if a valid response was given on over 50% of the items. All participants with valid information for at least one time point were included in the analyses.

3.1.2 Analytical procedure

The TRAILS data were analyzed with LCA, LCGA and GMM. A linear, quadratic and cubic slope were defined for LCGA and GMM. For GMM, if intercept or slope variances were estimated to be negative or very close to 0, they were set to 0. The analyses started with a model for one class and increased until the fit indices did not indicate further improvement or no optimal solution could be found. The optimal number of classes was determined by the fit indices mentioned for the simulation study, interpretability of the classes and the results of the simulation study. Morin et al indicate that for a sufficiently large sample size fit indices always favor the model with more classes (Morin et al. 2011). However, they do not specify what sufficiently large entails). The random intercept and slope variances and residual variances were allowed to differ across classes and over time. Whenever possible, analyses were reported according to the Guidelines for Reporting on Latent Trajectory Studies (GRoLTS) (van de Schoot et al. 2017).

3.2 Results

In TRAILS (Table 3), the results were highly similar for LCA, LCGA and GMM. The same types of classes were found, for the 1-class up to the 4-class solution (Supplementary Material 4). Table 4 shows the fit indices and entropy for LCA, LCGA and GMM for 1 up to 6 classes and Supplementary Material 4 shows the accompanying plots. For LCA and GMM, a stable solution could not be reached for 6 classes and more. For LCGA, this was the case for 7 classes and more. The BIC, aBIC and AIC kept decreasing as the number of classes increased. Based on these fit indices, the5-class model for LCA and GMM and the 6-class model for LCGA would be selected. The LMR-LRT indicated a 2-class solution for GMM and a 3-classs solution for LCA and LCGA.

In the 5-class solution, for LCGA and GMM, three different ‘decreasing’ classes emerged, with intercepts of 1.1, 2.6 and 4.6. These classes could be labeled “Very low and decreasing”, “Low and decreasing” and “Decreasing”. Since the first two classes were very similar, this solution was deemed less clinically relevant than the 4-class solution, where a “Low and decreasing” and “Decreasing” class emerged. Moreover, the decreases in BIC, aBIC and AIC from the 4-class to the 5-class model were not as substantial as the decreases in fit indices for e.g. the 2- versus 3-class model. Therefore, the 4-class model was selected. The 5-class LCA solution showed two very low classes, one decreasing and one increasing. Since the slope of both classes was very close to zero, the differences between these classes were also not considered to be clinically relevant. Therefore, for all methods, the 4-class solution was selected as the final model (Figure 2)

In the 4-class solution the first class was labeled “decreasing” and contained 30.1%, 30.0% and 31.8% of the TRAILS individuals for LCA, LCGA and GMM, respectively. The second class was labeled “low decreasing” (23.6% LCA, 23.8% LCGA, 22.9% GMM) was characterized by very little variation within the class. The third class was labeled “stable” (19.8% LCA, 19.9% LCGA, 25.5% GMM) and was characterized by high variation within the class. The fourth class was labeled “increasing” with class sizes of 26.0% (LCA), 25.9% (LCGA) or 19.9% (GMM) (Table 4). Comparison of the 4-class solution for LCA, LCGA and GMM shows no major differences in terms of background characteristics (Table 5). ”

Table 3 Background Characteristics in the Tracking Adolescent Individual Lives' Survey (TRAILS), The Netherlands, 2001-2017

Variables	N (%)	N
Sex, N (%) measured at age 11.1 (SD = 0.56)		2230
Male	1098 (49.2%)	
Female	1132 (50.8%)	
Parental education level measured at age 19.8 (SD = 0.60)		1677
Elementary education	25 (1.5%)	
Lower tracks secondary education	318 (19.0%)	
Higher tracks secondary education	539 (32.1%)	
Senior vocational education	422 (25.2%)	
University	373 (22.2%)	
Physical health, N (%) measured at age 11.1 (SD = 0.56)		2190
Poor	17 (0.8%)	
Acceptable	70 (3.2%)	
Fair	415 (18.9%)	
Good	1151 (52.6%)	
Excellent	537 (24.5%)	
Anxiety/depression	Mean (SD)	
Youth Self-Report at age 11.1 (SD = 0.56)	3.25 (2.72)	2195
Youth Self-Report at age 13.6 (SD = 0.53)	3.13 (2.92)	2093
Youth Self-Report at age 16.3 (SD = 0.71)	2.92 (2.30)	1660
Adult Self-Report at age 19.8 (SD = 0.60)	3.19 (3.28)	1694
Adult Self-Report at age 22.3 (SD = 0.65)	3.12 (3.41)	1499
Adult Self-Report at age 25.7 (SD = 0.60)	3.94 (3.79)	1316

N Number, *SD* Standard Deviation

Table 4 Model Fit Information for Latent Class Analysis (LCA), Latent Class Growth Analysis (LCGA) and Growth Mixture Modeling (GMM) of Anxiety/Depression Symptoms in the Tracking Adolescent Individual Lives' Survey (TRAILS), The Netherlands, 2001-2017

								Relative class sizes					
	Classes	BIC	aBIC	AIC	LMR-LRT p-value	Entropy	Parameters	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
LCA	1	53376	53338	53308	-	-	12	1.00					
	2	48295	48215	48152	0.00	0.80	25	0.51	0.49				
	3	46986	46866	46769	0.00	0.76	38	0.44	0.29	0.27			
	4	46566	46404	46275	0.14	0.70	51	0.31	0.27	0.21	0.21		
	5	46185	45982	45820	0.09	0.70	64	0.27	0.23	0.20	0.17	0.13	
	6[a]	-	-	-	-	-							
LCGA	1	53370	53338	53312	-	-	10	1.00					
	2	48285	48219	48165	0.00	0.80	21	0.49	0.51				
	3	46966	46865	46784	0.00	0.75	32	0.44	0.29	0.27			
	4	46539	46402	46293	0.14	0.70	43	0.31	0.27	0.21	0.21		
	5	46238	46066	45930	0.07	0.69	54	0.28	0.24	0.17	0.17	0.14	
	6	46060	45853	45688	0.29	0.66	65	0.21	0.21	0.17	0.15	0.12	0.13
GMM	1	50069	50006	49955	-	-	20	1.00					
	2	47331	47220	47131	0.00	0.74	35	0.55	0.45				
	3	46665	46525	46414	0.22	0.70	44	0.40	0.39	0.21			
	4	46350	46165	46018	0.05	0.66	58	0.31	0.28	0.23	0.19		
	5	46145	45945	45786	0.34	0.65	63	0.26	0.23	0.19	0.18	0.15	
	6[b]	46521	46273	46076	0.74	0.77	78	0.00	0.31	0.27	0.00	0.19	0.23

AIC Akaike Information Criterion, *aBIC* sample size adjusted BIC, *BIC* Bayesian Information Criterion, *LMR-LRT* Lo-Mendell-Rubin Likelihood Ratio Test
a) No optimal 6-class solution could be reached for LCA.
b) An improper solution was found for 6-class GMM, with 2 empty classes.

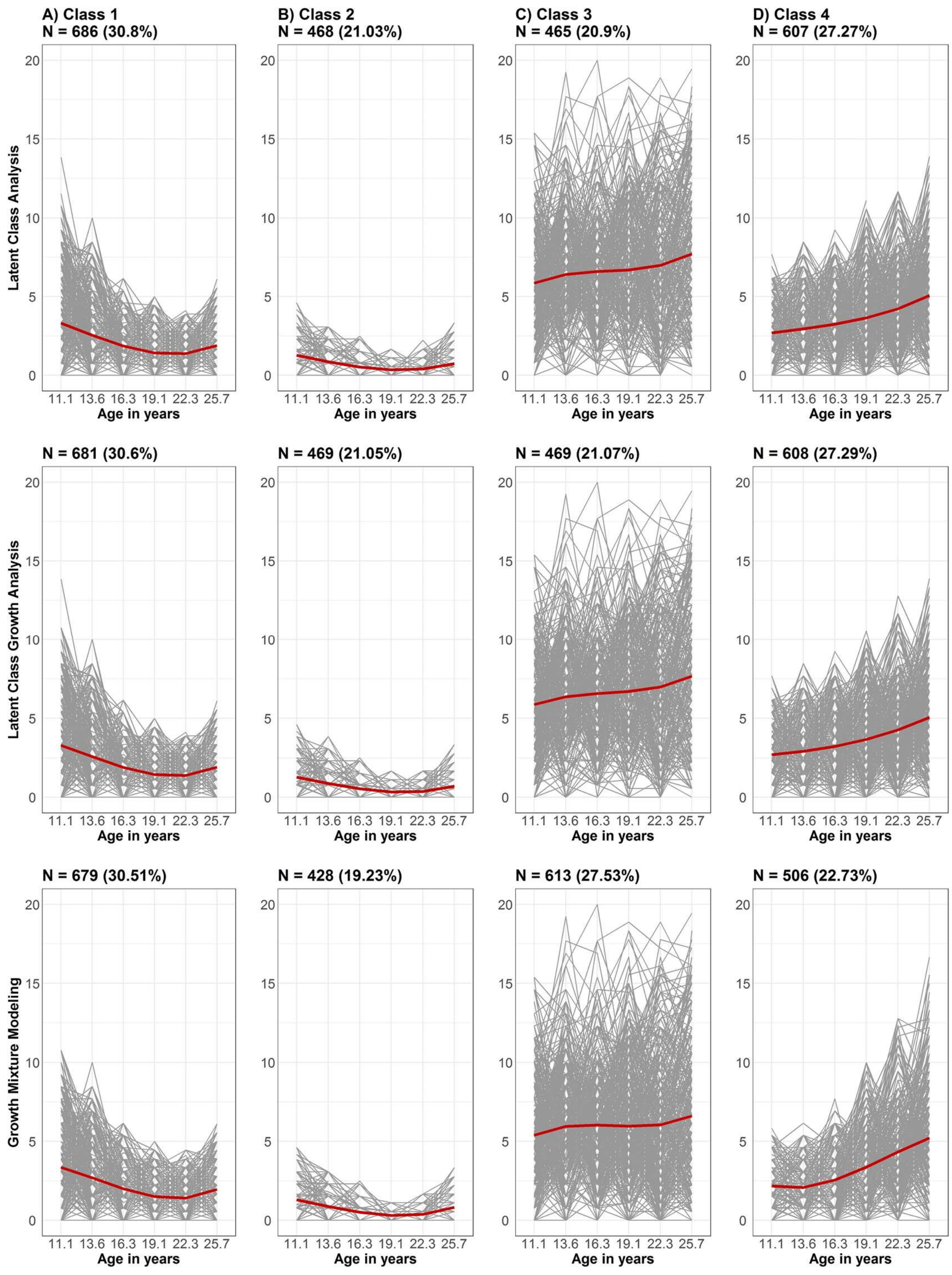

Figure 2 4-class Solutions for Anxiety/Depression Symptoms in the Tracking Adolescent Individual Lives' Survey (TRAILS) for Latent Class Analysis, Latent Class Growth Analysis and Growth Mixture Modeling. Note: The thick line represents the estimated mean trajectory for each class. The thin grey lines represent the observed trajectories of all individuals, with individuals assigned to classes based on their highest posterior probability to belong to a certain class.

Table 5 Background characteristics for the 4-class solution for latent class analysis, latent class growth analysis and growth mixture modeling of anxiety/depression symptoms in the Tracking Adolescent Individual Lives' Survey (TRAILS), The Netherlands, 2001-2017.

		LCA	LCGA	GMM
Class 1 "Decreasing"	Male, N (%)	343.5 (50.1%)	339.7 (49.9%)	334.3 (49.2%)
	Parents divorced, N (%)	142.3 (20.8%)	141.4 (20.8%)	142.5 (21.0%)
	Physical health T1, mean (SD)	3.9 (.82)	3.9 (.82)	3.9 (.82)
	Total class size, N (%)	685.9 (30.8%)	681.4 (30.6%)	679.4 (30.5%)
Class 2 "Low and Decreasing"	Male, N (%)	226.9 (48.4%)	227.8 (48.6%)	208.3 (48.6%)
	Parents divorced, N (%)	102.4 (21.9%)	102.2 (21.8%)	92.1 (21.5%)
	Physical health T1, mean (SD)	4.0 (.79)	4.0 (.78)	4.0 (.79)
	Total class size, N (%)	468.4 (21.0%)	468.7 (21.1%)	428.2 (19.2%)
Class 3 "Stable"	Male, N (%)	229.9 (49.4%)	231.9 (49.4%)	299.9 (48.9%)
	Parents divorced, N (%)	96.6 (20.8%)	97.3 (20.7%)	124.4 (20.3%)
	Physical health T1, mean (SD)	4.0 (.76)	4.0 (.76)	4.0 (.76)
	Total class size, N (%)	465.4 (20.9%)	469.1 (20.1%)	613.2 (27.5%)
Class 4 "Increasing"	Male, N (%)	282.8 (46.6%)	283.7 (46.7%)	240.8 (47.6%)
	Parents divorced, N (%)	116.0 (19.1%)	116.4 (19.2%)	98.3 (19.4%)
	Physical health T1, mean (SD)	4.0 (.78)	4.0 (.79)	4.0 (.79)
	Total class size, N (%)	607.3 (27.3%)	607.7 (27.3%)	506.1 (22.7%)

GMM Growth Mixture Modeling, *LCA* Latent Class Analysis, *LCGA* Latent Class Growth Analysis, *SD* Standard Deviation, *SES* Social Economical Status, *T1* Time point 1

4. Discussion

To our knowledge, this is the first study to compare LCA, LCGA and GMM to identify developmental trajectory classes using simulation data and a real data example, and which focused on bias. We found that GMM outperformed LCA and LCGA in terms of class recovery, bias, coverage and extraction of the simulated number of classes**,** if the data were generated according to a GMM The SE between replications was similar for the three methods. This result held under varying scenarios regarding sample size, number of time points, number of classes and relative class size. In four scenarios, the three methods performed comparably in terms of class recovery, namely a sample size of 8000, four time points, three underlying latent classes, large separation between classes and more equal class sizes. Only a large separation between classes influenced the relative model performance in terms of bias, resulting in equal performance of all methods. The analyses on the real data yielded rather similar estimated models for LCA, LCGA and GMM.

In the present study, GMM outperformed LCA and LCGA in terms of class recovery, bias, coverage and extraction of the simulated number of classes, under the condition that the data was generated according to a GMM. Previous studies which compared GMM and LCGA support our findings (Diallo et al. 2016; Davies et al. 2017). In contrast, Martin and von Oertzen found GMM to better identify the number of simulated classes than LCA for sample sizes below 600, but LCA was better for larger sample sizes. It remains unclear why LCA outperformed GMM for larger sample sizes in their study. Twisk and Hoekstra (Twisk and Hoekstra 2012) found that LCGA outperformed LCA and GMM for linear growth and that none of the methods performed well for quadratic growth. These inconsistencies may be due to differences in methodology. Twisk and Hoekstra manipulated an empirical dataset to consist of four latent classes, by standardizing it and adding a different number of SDs to the intercepts and slopes for different classes. Green (Green 2014) argued that the poor performance for quadratic growth was caused by residual population heterogeneity, that was picked up by the methods, and therefore the classes the authors aimed to manipulate into the data were not found. Standardization of methodology by simulation indeed shows GMM to outperform the other two methods.

We found GMM to be less biased than LCA and LCGA, regardless of sample size, number of time points, number of classes and relative class size. Kim (Kim 2012) found the minimal required sample size for GMM to depend strongly on the number of time points. Loughran and Nagin (Loughran and Nagin 2006) assessed the performance of LCGA for sample sizes of 500 up to 1,500 and found LCGA to perform sufficient for all sample sizes. Both Kim, and Loughran and Nagin only looked at the absolute performance of one method. We considered the relative performance of three methods and found GMM to outperform LCA and LCGA for sample sizes as small as 200 and still yield acceptable results.

Although GMM was less biased, and the simulated number of classes were extracted more often, substantial warnings during the data analyses occurred also most frequently for GMM. The most common warnings were due to negative variances. For the LCGA generated data this makes sense: the within class variation is 0 and is often estimated just below 0. However, these warnings were also common for the GMM generated data. In practice, GMM requires more tweaking than LCA and LCGA, for instance by setting the within group slope variance to 0, if estimated to be negative. Regardless of these warnings, the bias and class recovery for GMM was still better than for LCA and LCGA, but these warnings do require further attention.

Surprisingly, for seven time points, the class recovery was around 2% for LCA and LCGA, because the classes were severely distorted. In other words, after five time points, Class 1 and 3, and Class 2 and 4 partly overlapped. As a result, LCA and LCGA estimated classes that represent a mixture of the simulated classes. However, GMM accommodates for the variation within classes and could therefore still identify the simulated classes in most replications. Another surprising result was the lower bias in class 3 for a sample size of 200 for LCA and LCGA. A possible explanation is that for a sample size of 200, class recovery was successful in less than half of the replications. Therefore, less replications were included in the bias calculations, which artificially decreased the bias.

The real data example on anxiety/depression symptoms in TRAILS showed rather similar outcomes for LCA, LCGA and GMM. However, the class sizes did differ over the three methods, indicating that individuals were differently classified over the four classes by the three methods. These results confirm those of the simulation scenario that mimics the TRAILS data most closely: the scenario with more equal class sizes, with the largest class containing 40% of individuals. In the scenario with more equal class sizes, the class recovery was close to 100% for all methods (ranging from 99.3-100%) and the bias was small for all methods. The current study has several strengths. First, the systematic literature search provided input values for the simulation study, which increased the validity of our study. Second, we used a series of outcomes to determine which method was most appropriate (Burton et al. 2006). Third, the relative performance of the three methods was evaluated in different simulation scenarios to assess their influence on performance of the methods. Fourth, we combined a simulation study with a real data example, which allowed us to translate the implications of the simulation study to real data.

Several limitations need to be considered as well. First, we could not assess longitudinal measurement invariance in the real data example due to the use of two age-specific questionnaires with different items. This is unlikely to have affected our findings to a large extent, as the three methods can be expected to be similarly sensitive. Second, in a simulation study data has to be generated according to a specific model. We generated according to two different models, LCGA and GMM. Therefore, we could compare the results for both data generations. We deem GMM to be most realistic, since it accounts for more variation within classes. Real data might be somewhere in between LCGA and GMM.

Our findings imply that further research is needed on several issues. It remains unclear how the three methods perform if model assumptions are not met, for example, if the distribution within classes is non-normal. We only investigated the methods' abilities to estimate developmental trajectories. Future research might explore the performance of LCA, LCGA and GMM in the presence of covariates, for instance when using the classes to predict observed outcomes (Bakk et al. 2013) or other trajectories (Nagin and Odgers 2010; Nagin et al. 2016) or when using the classes as an outcome (Vermunt 2010).

4.1 Conclusion

In conclusion, our findings suggest that researchers who aim to identify developmental trajectory classes should prefer GMM over LCA and LCGA, if variations around the intercept or slope parameters are substantial. This advice holds irrespective of sample size, number of time points, expected number of classes, expected separation between classes and expected class sizes. If variations around the intercept and slope parameters are substantial within classes, GMM can be expected to be least biased. If there is no variation around the class intercept or slope within classes, all three methods can be expected to be equally unbiased. To summarize, GMM seems to be the most appropriate method to identify developmental trajectory classes under various simulated conditions.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

Achenbach T, Rescorla L (2003) Manual for the ASEBA Adult Forms & Profiles. Burlington

Achenback TM, Rescorla L a. (2001) Manual for the ASEBA School-Age Forms & Profiles

Bakk Z, Tekle FB, Vermunt JK (2013) Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. Sociol Methodol 43:272–311. doi: 10.1177/0081175012470644

Berlin KS, Parra GR, Williams NA (2014a) An introduction to latent variable mixture modeling (Part 2): Longitudinal latent class growth analysis and growth mixture models. J Pediatr Psychol 39:188–203. doi: 10.1093/jpepsy/jst085

Berlin KS, Williams NA, Parra GR (2014b) An introduction to latent variable mixture modeling (Part 1): Overview and cross-sectional latent class and latent profile analyses. J Pediatr Psychol. 39:174–187

Burton A, Altman DG, Royston P, Holder RL (2006) The design of simulation studies in medical statistics. Stat Med 25:4279–4292. doi: 10.1002/sim.2673

Chen F, Bollen KA, Paxton P (2001) Improper Solutions in Structural Equation Models Causes, Consequences, and Strategies. Sociol Methods Res 29:468–508

Davies CE, Glonek GF, Giles LC (2017) The impact of covariance misspecification in group-based trajectory models for longitudinal data with non-stationary covariance structure. Stat Methods Med Res 26:1982–1991. doi: 10.1177/0962280215598806

Daw J, Margolis R, Wright L (2017) Emerging Adulthood, Emergent Health Lifestyles: Sociodemographic Determinants of Trajectories of Smoking, Binge Drinking, Obesity, and Sedentary Behavior. J Health Soc Behav 58:181–197. doi: 10.1177/0022146517702421

Diallo TMO, Morin AJS, Lu H (2016) Impact of Misspecifications of the Latent Variance–Covariance and Residual Matrices on the Class Enumeration Accuracy of Growth Mixture Models. Struct Equ Model A Multidiscip J 23:507–531. doi: 10.1080/10705511.2016.1169188

Feldman BJ, Masyn KE, Conger RD (2009) New approaches to studying problem behaviors: a comparison of methods for modeling longitudinal, categorical adolescent drinking data. Dev Psychol 45:652–676. doi: 10.1037/a0014851

Gangestad S, Snyder M (1985) " To Carve Nature at Its Joints ": On the Existence of Discrete Classes in Personality. Psychol Review 92:317–349

Green MJ (2014) Latent class analysis was accurate but sensitive in data simulations. J Clin Epidemiol 67:1157–1162. doi: 10.1016/j.jclinepi.2014.05.005

Hallquist MN, Wiley JF (2018) MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. Struct Equ Model. 1–18

Hoekstra T (2013) Applied latent class models for epidemiology Dissertation, VU University Amsterdam

Huisman M, Oldehinkel AJ, de Winter A, et al (2008) Cohort profile: the Dutch “TRacking Adolescents” Individual Lives’ Survey’; TRAILS. Int J Epi 37:1227–1235. doi: 10.1093/ije/dym273

Kim S-Y (2012) Sample Size Requirements in Single- and Multiphase Growth Mixture Models: A Monte Carlo Simulation Study. Struct Equ Model 19:457–476. doi: 10.1080/10705511.2012.687672

Kooken J, McCoach DB, Chafouleas SM (2018) The Impact and Interpretation of Modeling Residual Noninvariance in Growth-Mixture Models. J Exp Educ. 1–24

Lanza ST, Collins LM (2006) A mixture model of discontinuous development in heavy drinking from ages 18 to 30: the role of college enrollment. J Stud Alcohol 67:552–561

Lo Y, Mendell NR, Rubin DB (2001) Testing the number of components in a normal mixture.

Biometrika 88:767–778

Loughran T, Nagin DS (2006) Finite sample effects in group-based trajectory models. Sociol Methods Res 35:250–278. doi: 10.1177/0049124106292292

Martin DP, von Oertzen T (2015) Growth Mixture Models Outperform Simpler Clustering Algorithms When Detecting Longitudinal Heterogeneity, Even With Small Sample Sizes. Struct Equ Model 22:264–275. doi: 10.1080/10705511.2014.936340

Morin AJS, Maïano C, Nagengast B, et al (2011) General growth mixture analysis of adolescents' developmental trajectories of anxiety: The impact of untested invariance assumptions on substantive interpretations. Struct Equ Model 18:613–648. doi: 10.1080/10705511.2011.607714

Muthen B, Shedden K (1999) Finite Mixture Modeling with Mixture Outcomes Using the EM Algorithm. Biometrics 55:463–469

Muthén LK, Muthén BO Mplus User's Guide. Muthén & Muthén, Los Angeles, CA

Nagin DS (1999) Analyzing developmental trajectories: A semiparametric, group-based approach. Psychol Methods 4:139

Nagin DS, Jones BL, Lima Passos V, Tremblay RE (2016) Group-based multi-trajectory modeling. Stat Methods Med Res 096228021667308. doi: 10.1177/0962280216673085

Nagin DS, Odgers CL (2010) Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 6:109–38. doi: 10.1146/annurev.clinpsy.121208.131413

Nagin DS, Tremblay RE (2005) Developments Trajectory Groups: Fact or a Useful Statistical Fiction. Criminology 43:873–904

Oldehinkel AJ, Rosmalen JG, Buitelaar JK, et al (2015) Cohort profile update: The Tracking Adolescents' Individual Lives Survey (TRAILS). Int J Epidemiol 44:76–76n. doi:

10.1093/ije/dyu225

Patel M, Kaufman A, Hunt Y, Nebeling L (2017) Understanding the Relationship of Cigarette Smoking Trajectories Through Adolescence and Weight Status in Young Adulthood in the United States. J Adolesc Heal 61:163–170. doi: 10.1016/j.jadohealth.2017.02.005

Paxton P, Curran PJ, Bollen KA, et al (2001) Monte Carlo Experiments: Design and Implementation. Struct Equ Model A Multidiscip J 8:287–312. doi: https://doi.org/10.1207/S15328007SEM0802_7

R Core Team (2017) R: A Language and Environment for Statistical Computing. https://www.r-project.org

Sher KJ, Jackson KM, Steinley D (2012) Alcohol Use Trajectories and the Ubiquitous Cat's Cradle : Cause. 120:322–335. doi: 10.1037/a0021813.Alcohol

Twisk J (2013) Applied Longitudinal Data Analysis for Epidemiology: A Practical Guide., 2nd editio. Cambridge University Press

Twisk J, Hoekstra T (2012) Classifying developmental trajectories over time should be done with great caution: A comparison between methods. J Clin Epidemiol 65:1078–1087. doi: 10.1016/j.jclinepi.2012.04.010

van de Schoot R, Sijbrandij M, Winter SD, et al (2017) The GRoLTS-Checklist: Guidelines for Reporting on Latent Trajectory Studies. Struct Equ Model 00:1–17. doi: 10.1080/10705511.2016.1247646

Veldman K, Reijneveld SA, Verhulst FC, et al (2017) A life course perspective on mental health problems, employment, and work outcomes. Scand J Work Environ Health 43:316–325. doi: 10.5271/sjweh.3651

Vermunt JK (2010) Latent class modeling with covariates: Two improved three-step approaches. Polit Anal 18:450–469. doi: 10.1093/pan/mpq025

Virtanen P, Nummi T, Lintonen T, et al (2015) Mental health in adolescence as determinant of alcohol consumption trajectories in the Northern Swedish Cohort. Int J Public Health 60:335–342. doi: 10.1007/s00038-015-0651-5

Wickrama KAS, Conger RD, Todd Abraham W (2008) Early family adversity, youth depressive symptom trajectories, and young adult socioeconomic attainment: A latent trajectory class analysis. Adv. Life Course Res. 13:161–192